

Hidden “Digestome”: Current Analytical Approaches Provide Incomplete Peptide Inventories of Food Digests

Maristella De Cicco,[†] Gianfranco Mamone,[†] Luigia Di Stasio,^{†,‡} Pasquale Ferranti,[‡] Francesco Addeo,[‡] and Gianluca Picariello^{*,†,‡}

[†]Institute of Food Sciences, National Research Council (CNR), Via Roma 64, 83100 Avellino, Italy

[‡]Department of Agriculture, University of Naples “Federico II”, Parco Gussone, Via Università 100, 80055 Portici, Naples, Italy

S Supporting Information

ABSTRACT: Analyzing an *in vitro* gastroduodenal digest of whey proteins by high-performance liquid chromatography (HPLC) coupled to high-resolution/high-sensitivity tandem mass spectrometry (MS/MS), we sought to evaluate if state-of-art peptidomics provide comprehensive peptide coverage of food “digestomes”. A multitude of small-sized peptides derived from both α -lactalbumin and β -lactoglobulin as well as disulfide cross-linked hetero-oligomers remained unassigned, even when the digests were compared before and after S–S reduction. The precipitation with 12% trichloroacetic acid demonstrated the occurrence of large-sized polypeptides that escaped the bioinformatic identification. The analysis of a HPLC–MS/MS run with different proteomic search engines generated dissimilar peptide subsets, thus emphasizing the demand of refined searching algorithms. Although the MS/MS fragmentation of monocharged ions with exclusion of non-peptide-interfering compounds enlarged the inventory of short peptides, the overall picture of the “digestome” was still incomplete. These findings raise relevant implications for the identification of possible food-derived bioactive peptides or allergenic determinants.

KEYWORDS: food peptide digestomes, bioactive peptides, food allergens, HPLC–MS/MS, disulfide cross-linked peptides, protein digestibility, whey proteins

INTRODUCTION

The research focused on discovery of food-derived bioactive peptides, and assessment of their activity has increased exponentially over the last few decades. To date, bioactive peptides have been described in a very ample range of food matrices, including milk and dairy products, eggs, meat, fish, legumes, and cereals. Bioactive peptides are linear amino acid sequences that remain inactive as long as they are encrypted in the parent protein and became active upon proteolytic release, exhibiting a wide range of more or less evident effects *in vivo*.¹ The gastrointestinal (GI) stability of food-derived peptides is an obvious requirement to exert bioactive effects, at either a systemic or local level.

Recently, the release of possible bioactive peptides from foods has been predicted with bioinformatic approaches on the basis of known proteolytic specificity of digestive proteases, expected stability of protein domains, and interaction models of bioactive motifs with receptors, enzymes, or antibodies.² Combinatorial models, computational advances, and combined analytical workflows concur to predict the formation of peptide components within complex food-derived digests.^{3,4} However, in general, *in silico* predictions do not mirror the *in vitro* or *in vivo* outcomes as a consequence of the molecular and structural complexity of the food matrices, process-induced modifications, and inter-/intraindividual variability of the human digestion physiology. Thus, the typical pipeline to discover single or panels of candidate peptides able to induce a response in relevant biological systems (molecular targets, cells, tissues, and organisms) is still firmly ground on the peptidomic characterization of food digests obtained through *in vitro*, *ex*

vivo, or *in vivo* models. Several strategies can be used to enrich/isolate and characterize specific peptides from complex digests, to test them in biological assays.⁵ The most advanced methods in this sense make extensive use of mass spectrometry (MS).

Generally, protein digestomes comprise free amino acids, small-/medium-size peptides, large polypeptides, also including possible disulfide-linked heterodimers/oligomers, and in some cases variable amounts of undigested proteins. Ordinary proteomic/peptidomic workflows involve automated ion selection for tandem mass spectrometry (MS/MS) fragmentation (data-dependent acquisition), excluding singly charged species to limit the interference by background low-molecular-weight compounds.

As a result of the complexity of the digestomes, the eventuality of assigning hundreds to thousands individual peptides manually is impracticable. Therefore, more than 30 different search engines have been developed in the last few decades to identify peptides in complex proteolytic mixtures.⁶ Some of them perform spectral library searching or are capable of *de novo* sequencing, whereas others match experimentally obtained fragmentation spectra with theoretical spectra generated from a protein sequence database.⁷ In all cases, the matching score from available algorithms increases proportionally with the peptide length, thus intrinsically underrepresenting small peptides, which often comprise

Received: April 14, 2019

Revised: May 14, 2019

Accepted: May 15, 2019

Published: May 15, 2019

those involved in bioactive actions. The identification of very short peptides (<5 amino acids) could also be hindered by insufficient fragmentation patterns, inadequate to distinguish between several possible isobaric sequences.⁸ Thus, the untargeted identification of small-sized peptides in a food digestome can severely challenge the potentiality of peptidomics.⁹

Several expedients have been developed to overcome these limitations, including (i) N-terminal derivatization,¹⁰ (ii) prediction methods of hydrophobicity and retention time coupled to hydrophilic interaction liquid chromatography (HILIC)–MS/MS,⁸ and (iii) prediction of the released peptides and reversed-phase ultra-performance liquid chromatography (UPLC)–MS/MS combined to a *de novo* sequencing stage.¹¹ These methods have advantages and disadvantages; however, they all introduce more or less laborious steps in the identification process, and for this reason, they are not yet routinely used.

The MS detection of food-derived large-sized polypeptides in a complex digest can be problematic as well, due to suppression ion effects, reduced sensitivity associated with wide charge distribution, and limited fragmentation efficiency.¹² A number of structural features, such as low solubility, intrinsically scarce ionization efficiency, presence of post-translational modifications (e.g., multiple phosphorylation or glycosylation), incompleteness of the protein databases, variability of protein isoforms, and presence of hard to identify disulfide cross-linked peptide heterodimers,^{13,14} further limits the comprehensive characterization of a food digestome. The complexity can even increase when foods undergo thermal processing as a result of variable protein modifications.

The intrinsic stability to GI digestion of large domains has long been considered among the primary factors in the allergenicity risk assessment of food proteins.¹⁵ Many food allergens are proteins highly resistant to proteolytic enzymes. Likewise, the GI stability of gluten proteins has a recognized role in the pathological mechanisms of celiac disease.^{16,17} For this reason, identification of food protein sequences surviving GI digestion is largely used to predict possible peptide determinants of adverse reaction to foods. On the other hand, the correlation of digestion stability with sensitizing/eliciting capacity in the etiopathology of food allergy still appears controversial. There are more than a few examples of food allergens, such as α -lactalbumin, caseins, thermally treated ovalbumin, and peanut (*Arachis hypogaea*) Ara h 1,¹⁸ which are believed to be substantially labile and susceptible to fast digestion proteolysis. In effects, these proteins are broken down into small peptides within seconds or minutes when incubated with pepsin or downstream with duodenal and intestinal proteases/peptidases. Nevertheless, the survival of potentially immunogenic/allergenic large polypeptides at low abundance has not been adequately investigated.^{19,20} The minimal dose of determinants able to sensitize or elicit an allergic response is not known, and in principle, even very small amounts of polypeptides that survived digestion could be responsible for an allergic response in predisposed individuals. These aspects are still underexplored, because stability or susceptibility to digestion of food proteins has been typically evaluated monitoring the disappearance of the protein bands by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS–PAGE) analysis, while only sporadically, the peptide digests have been characterized in depth using MS-based peptidomics. Even using advanced techniques, the character-

ization of a complex “digestome” is a challenging task, which heavily depends upon the fractionation and enrichment of a specific set of components as well as technical and contingent performances of MS instrumental resources.²¹ Notably, standard procedures for a comprehensive characterization of food digests do not yet exist.

Herein, through the high-resolution high-performance liquid chromatography (HPLC)–MS/MS characterization of a digest of whey proteins obtained with a static *in vitro* method of simulated digestion,²² we tried to address the non-trivial question: do the current analytical approaches provide comprehensive peptide inventory of a food digest?

This issue has relevant implications for both discovery of possible food-derived bioactive peptides and evaluation of stability/susceptibility of food proteins to digestion, in view of the allergenicity risk assessment.

MATERIALS AND METHODS

All chemicals were purchased from Sigma–Aldrich (St. Louis, MO, U.S.A.) and were analytical-grade or better. HPLC-grade solvents were purchased from Carlo Erba (Milan, Italy).

Isolation of Whey Proteins. Whey proteins were chosen as a relatively simple model of food proteins. Bulk bovine milk was obtained from a local farm and collected soon after milking. Milk samples (50 mL) were aliquoted into sterile polystyrene containers. To prevent undesired proteolysis, phenylmethylsulfonyl fluoride (PMSF) was added to a final concentration of 1 mM and milk was frozen and stored at -20°C until use. Milk was centrifuged at 1500g for 20 min at 4°C (Labofuge 400R, Heraeus Instruments, Hanau, Germany) and skimmed by manual removal of the floating fat layer (twice). Casein was depleted by isoelectric precipitation at pH 4.6 with 10% (v/v) acetic acid and 1 M sodium acetate and pelleted by centrifugation at 1500g for 10 min at 4°C . Whey proteins were purified from the supernatants dialyzing against deionized water, using benzoylated dialysis tubes with 3 kDa molecular weight cutoff (Sigma).

Afterward, proteins were quantified with the Bradford assay (kit from Bio-Rad, Milan, Italy) and finally lyophilized.

It was verified by reversed-phase high-performance liquid chromatography with ultraviolet detection (RP-HPLC–UV) analysis that whey proteins contained both the A and B of β -lactoglobulin (β -Lg), using the analysis conditions by Ferranti et al.²³

In Vitro Digestion of Whey Protein. *In vitro* simulated gastroduodenal (GD) digestion of whey proteins was performed according to Minekus et al.,²² skipping the oral phase. Simulated gastric fluid (SGF) and simulated intestinal fluid (SIF) were prepared according to the harmonized conditions. All digestion steps were carried out in a shaking incubator at 37°C and 170 rpm. Briefly, for the gastric phase, 500 mg of whey proteins was diluted with 400 μL of SGF 1.25X. Liposomes, freshly prepared with lecithin (Sigma), were added up to 0.17 mM in the final solution. The pH was adjusted to 2.7 and porcine pepsin (2000 units/mL final concentration). The sample was incubated for 2 h at 37°C . Pepsin hydrolysis was stopped by raising the pH to 7.0 with 1 M sodium bicarbonate. The duodenal digestion was carried out for 2 h at 37°C after duplicating the volume with SIF, containing bile salts (10 mM final on the basis of cholic acid) and pancreatin at 100 units/mL based on the trypsin activity [7 units/mg of pancreatin as *N*-4-tosyl-L-arginine methyl ester (TAME) activity]. At the end of the duodenal phase, peptides in a 1 mL aliquot of the digest were immediately purified with C18 reverse-phase pre-packed cartridges (Sep-Pak, Waters, Milford, MA, U.S.A.), washing extensively with 0.1% (v/v) trifluoroacetic acid (TFA) and eluting with 70% acetonitrile/0.1% TFA. Peptides were vacuum-dried and finally lyophilized. An aliquot of the peptide extract (10 μg) was suspended in 100 μL of 25 mM ammonium bicarbonate/10 mM dithiothreitol (DTT) and incubated at 55°C for 45 min prior to LC–MS/MS analysis. The remaining solution of digests was aliquoted, immediately frozen, and stored at -20°C until use. A solution of SGF

and SIF containing the respective digestive enzymes but lacking the whey proteins was incubated under the same conditions of the sample and used as the control.

Pre-fractionation of the Whey Protein Digests. Several attempts of pre-fractionation of the whey protein digest were performed, including size-exclusion chromatography (SEC) using either Econo-Pac (Bio-Rad, Hercules, CA, U.S.A.) pre-packed columns (exclusion limit of 6 kDa) or polyacrylamide spin desalting columns (Pierce–Thermo Scientific), in both cases eluting the high-molecular-weight fraction with 25 mM ammonium bicarbonate at pH 7.8. Alternatively, aliquots (1 mL) of the digests were diluted with 100% (w/v) trichloroacetic acid (TCA) up to 12% final TCA concentration. After 30 min in an ice-cold bath, the precipitate was pelleted by centrifugation (3000g for 15 min at 4 °C) and the supernatant was discarded. To remove the excess TCA, the pellet was washed thrice with –20 °C cold acetone and finally dried.

SDS–PAGE Analysis. The course of digestion was monitored by SDS–PAGE using hand-cast 15% acrilamide gels (Tetra Cell Mini-PROTEAN system, Bio-Rad). Indicatively, 20 µg of proteins, suspended in Laemmli buffer, was loaded onto each well of the gel. Proteins were separated at room temperature, at 100 V constant voltage, and were visualized with Coomassie Brilliant Blue G-250. The gel was imaged with a scanner and processed using the LABScan software 3.00 (Amersham Bioscience, Uppsala, Sweden).

RP-HPLC. RP-HPLC analysis of whey protein digests and the relevant 12% TCA-insoluble fraction was carried out using a HP 1100 HPLC chromatographer (Agilent, Palo Alto, CA, U.S.A.) equipped with a C18 250 × 2.0 mm inner diameter, 4 µm particle diameter column (Jupiter, Phenomenex, Torrance, CA, U.S.A.). After 5 min of isocratic elution at 5% solvent B (0.1% TFA in acetonitrile, v/v), 5–60% B over 60 min was applied at a flow rate of 0.2 mL/min. Solvent A was 0.1% TFA in water (v/v). The column was maintained in a thermostatic oven at 40 °C. The column effluents were monitored by UV detection at $\lambda = 220$ and 280 nm. Nearly 50 µg of peptide digests was analyzed for each run. In the case of the 12% TCA-insoluble fraction, the pellet resulting from a 1 mL aliquot of the digest was resuspended in 200 µL of 0.1% TFA and 50 µL was injected onto the column. Undigested whey proteins (15 µg) were separated under the same conditions of the digests, to confirm the digestion of the proteins and to check the possible permanence of intact protein chains.

Nanoflow HPLC–MS/MS. HPLC–MS/MS analyses were performed using Ultimate 3000 nanoflow ultra-high-performance liquid chromatography (Dionex/Thermo Scientific, San Jose, CA, U.S.A.) coupled to a Q-Exactive Orbitrap mass spectrometer (Thermo Scientific). Peptides were reconstituted in 0.1% formic acid, and nearly 2 µg was loaded through Acclaim PepMap 100 trap columns (75 µm inner diameter × 2 cm, Thermo Scientific) using a FAMOS autosampler (Thermo Scientific). Peptides were separated with an EASY-Spray PepMap C18 column (2 µm, 25 cm × 75 µm) with 2 µm particles and 100 Å pore size (Thermo Scientific), applying a 2–50% gradient of B over 120 min after 5 min of isocratic elution at 2% B and a constant flow rate of 300 nL/min. Eluent A was 0.1% formic acid (v/v) in LC–MS-grade water, and eluent B was 0.1% formic acid (v/v) in acetonitrile. MS1 precursor spectra were acquired in the positive ionization mode, scanning the m/z 1800–350 range with a resolving power of 70 000 full width at half maximum (fwhm), an automatic gain control (AGC) target of 1×10^6 ions, and maximum ion injection time (IT) of 256 ms. The spectrometer operated in data-dependent acquisition mode while selecting up to the 10 most intense ions for MS/MS fragmentation and applying a 12 s dynamic exclusion. Fragmentation spectra were obtained at a resolving power of 17 500 fwhm. Ions with one charge or more than six were excluded from the MS/MS selection.

To improve the identification of short peptides, the HPLC–MS/MS analysis of peptide digests, both non-reduced and reduced with DTT, was repeated under the same chromatographic conditions, scanning the m/z 1600–200 range. In this case, fragmentation of monocharged ions with intensity higher than 1×10^6 was also allowed, while selection of background ions was impeded with an

exclusion list, which was prepared on the basis of a blank HPLC–MS/MS analysis. Spectra were elaborated using the software Xcalibur, version 3.1 (Thermo Scientific).

Identification of Peptides. The LC–MS/MS runs were used for interrogating a home-built database of the 30 most abundant proteins of bovine milk,²⁴ using the following search engines: (i) the ProteinProspector Batch-Tag web tool (<http://prospector2.ucsf.edu>), (ii) the Andromeda tool of MaxQuant, version 1.6.2.10, and (iii) the Proteome Discoverer 2.1 (Thermo Scientific), relying on the SEQUEST algorithm. For the analysis with the Batch-Tag, the mgf files were generated from the LC–MS runs using the MS Convert tool of the open source ProteoWizard 3.0 software (<http://proteowizard.sourceforge.net/>). Database searching parameters were the following in all cases: no modification of cysteines, Met oxidation and pyroglutamic acid at the N terminus of Gln as variable modifications, mass tolerance value of 8 ppm for precursor ion and 12 ppm for MS/MS fragments, and no enzyme specificity for the cleavage. The LC–MS/MS analyses of the peptide digests reduced by DTT were processed considering the possibility that cysteines were in the free thiol form. Charges of precursor peptides were (+)2–5 or (+)1–5 for analysis carried out with the possibility of fragmenting monocharged ions.

The Batch-Tag web search engine was operated also considering the possibility of peptide cross-linking through disulfide bridges.

Identification scores were calculated by target-decoy peptide-spectrum match (PSM) filtering, working at 0.01 peptide-level false discovery rate (FDR). In the case of the Proteome Discoverer search engine, identifications were validated with the Percolator tool.

RESULTS AND DISCUSSION

Monitoring the Protein Degradation. The course of digestion of whey proteins was monitored by SDS–PAGE analysis (Figure 1A). In agreement with previous findings,²⁵ a

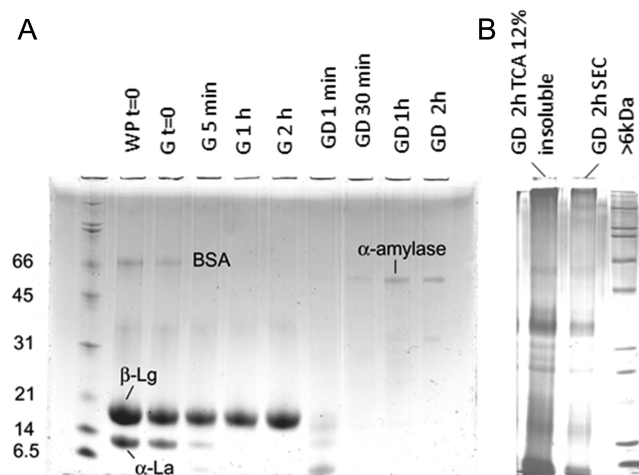


Figure 1. (A) SDS–PAGE monitoring of the *in vitro* simulated GD digestion of whey proteins. While β -Lg was partly resistant to simulated gastric (G) digestion, no whey protein-derived polypeptide was detected after the gastroduodenal (GD) step. (B) On the other hand, large polypeptides (MW range) were detected after selective enrichment of the digests by 12% TCA precipitation and SEC (Econo-Pac cartridge, with an exclusion limit of 6 kDa).

significant part of β -Lg resisted the simulated gastric digestion, while it was quickly degraded by simulated duodenal fluids.²⁶ In contrast, α -La and bovine serum albumin (BSA) were extensively digested already at the level of the simulated gastric compartment. After a few minutes of *in vitro* duodenal digestion, no polypeptides were detected by SDS–PAGE, apart from digestive enzymes. The volume of SDS–PAGE–

separated digests corresponded to 20 μg of the original protein content. With this protein amount, undigested proteins appeared slightly overloaded, whereas only faint bands of digestive enzymes were detected after complete digestion. The SDS–PAGE analysis of digests was repeated after concentration by SEC or precipitation with TCA (final concentration of 12%, w/w), demonstrating the presence of polypeptides with estimated molecular weights (MWs) of 4–6 and 15–16 kDa, which most likely originated from whey proteins (Figure 1B).

This finding suggests that food proteins appearing extensively digested could actually generate low abundance of large fragments, which became detectable upon concentration. At the moment, a minimum threshold amount for food allergens has not been established. Indeed, in principle, even scant amounts of polypeptides might sensitize and/or elicit an allergic reaction in predisposed individuals. Thus, the persistence of digestion-stable large polypeptides, albeit at very low abundance, raises important implications about the correlation between digestion stability and food allergenic potential. The possible presence of these protein fragments after digestion has been most often overlooked for a large variety of food allergens. The definitive identification of these large polypeptides would require dedicated time-consuming approaches, combining electrophoretic and/or HPLC isolation and MS analysis.

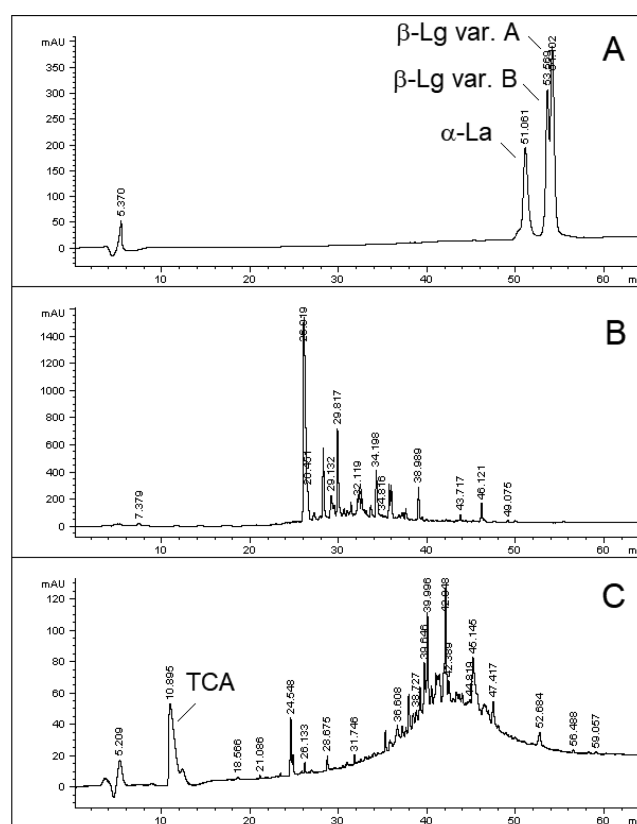
HPLC Analysis of Whey Protein Digests. As expected, the RP–HPLC analysis of GD digests (Figure 2B) showed the formation of a multitude of peptide fragments and confirmed the complete degradation of whey proteins (Figure 2A).

Interestingly, the peptide profiles of Cys non-reduced and Cys reduced by DTT exhibited only minor differences (not shown), suggesting that possible disulfide cross-linked peptides could occur at a low abundance or are less detectable than the linear peptide fragments. Large polypeptide fragments were selectively targeted, analyzing the 12% insoluble TCA fraction in the same HPLC conditions (Figure 2C). The relevant chromatogram contained a multitude of peaks eluting at a relatively high retention time, which were not evidenced in the analysis of unfractionated digests, probably as a consequence of suppression by dominant peptides. In line with SDS–PAGE analysis, these results suggested that the digest does contain large polypeptide fragments. At least part of these large-sized polypeptides arose from whey proteins rather than being impurities or hydrolysis products of digestion fluids, because they were substantially missing in the HPLC chromatogram of a control sample prepared with GD digestive enzymes in the absence of the whey protein substrate, incubated, and then fractionated under parallel conditions (not shown).

HPLC–MS/MS Identification of Whey Protein-Derived Peptides. To maximize the identification of the peptide components, we separated the whey protein digests by HPLC–MS/MS, with a 120 min gradient and using a high-resolution 25 cm capillary column.

The total ion current (TIC) chromatogram (Figure 3A) was largely dominated by β -Lg f(125–135) ($[M + 2H]^{2+} = 623.29$), which was previously demonstrated to be relatively resistant to simulated GI digestion,¹³ and β -Lg f(149–154) ($[M + H]^+ = 678.34$), eluting at retention times (t_r) of 48.62 and 61.45 min, respectively.

This latter peptide was tentatively identified only with the accurate MW; in fact, it was not fragmented because its possibly doubly charged ions were out of the range of analysis



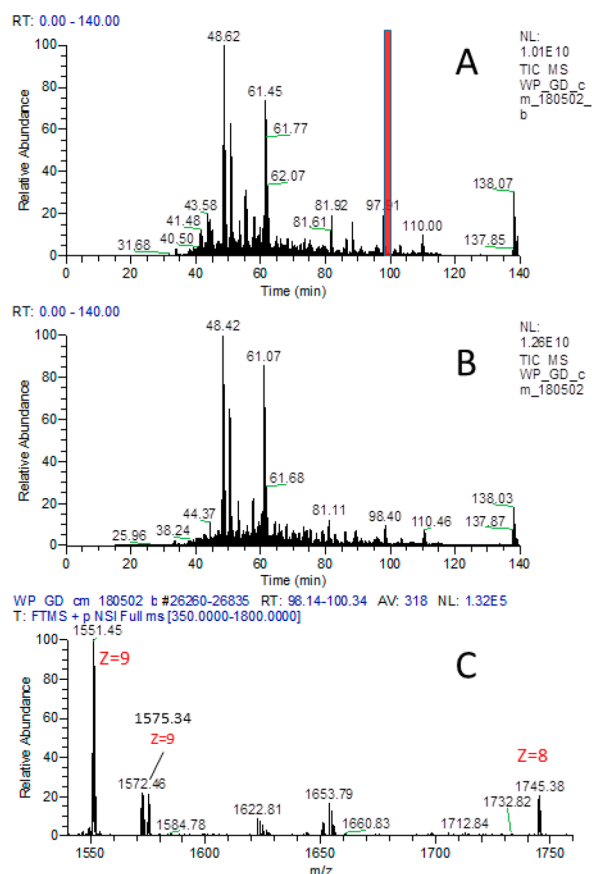


Figure 3. (A) TIC chromatogram of the HPLC–MS/MS analysis of *in vitro* gastrointestinal digests of whey proteins. (B) Reduction with DTT only slightly modified the TIC peptide profile, suggesting that cross-linked peptides might occur in low-intensity chromatographic peaks. (C) MS spectrum at a retention time of 98.1–100.0 min (highlighted in red in the TIC chromatogram) showed the presence of very low-abundant components with a MW of 13 954 ($[M + 9H]^+ = 1554.45$, and $[M + 8H]^+ = 1745.38$).

prevent the continuous fragmentation of interfering compounds, introduced for example with the eluents, the analysis was carried out with an opportune fragmentation exclusion list compiled through a blank HPLC–MS/MS acquisition.

Interestingly, this latter workflow confirmed the identification of dominant β -Lg f(149–154) by MS/MS sequencing. Similarly, assignment of BSA f(414–421) and α -La f(18–25) was confirmed, although they were retrieved by only two or one search engine, respectively.

Software-Assisted Identification of Peptide Digests.

The identification of proteins using HPLC–MS/MS-based “peptide-centric” approaches is a keystone in bottom-up proteomics, which for some aspects has not yet been completely addressed. In general, the confident assignment of peptides in proteolytic mixtures, especially in complex digests producing large data sets, depends upon the algorithm and the search engine performance. Using the same HPLC–MS/MS analysis with the same searching parameters, we tested three different search engines, namely, the open sources Andromeda tool of MaxQuant, the Batch-Tag of the ProteinProspector suite, and the commercial Proteome Discoverer 2.1 (Thermo Scientific), purchased with the spectrometer.

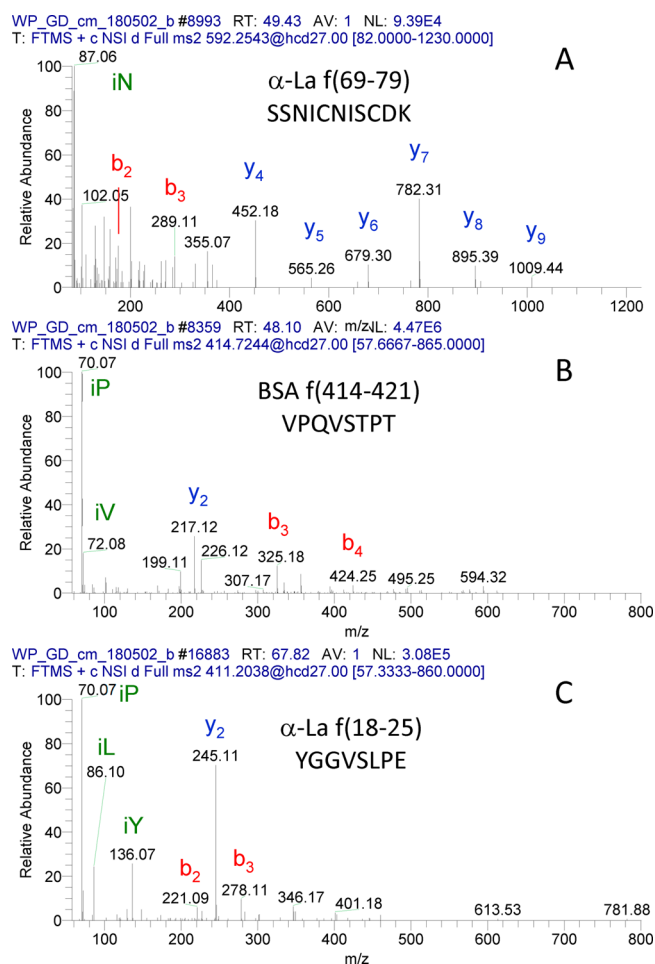


Figure 4. Manual assignment of MS/MS spectra to peptides escaped the identification by search engines: (A) α -La f(69–79), sequence: SSNICNISC DK; (B) BSA f(414–421), sequence: VPQVSTPT; and (C) α -La f(18–25) sequence: YGGVSLPE. Fragments of the b and y series are labeled in red and blue, respectively. Immonium ions are labeled in green.

Figure 5A shows a Venn diagram with sets of milk-derived peptides, identified from non-reduced digests by each of the search engines and their mutual intersections. The ProteinProspector and the Proteome Discoverer exhibited the best and worst performances, with overall 116 and 45 identified peptides, respectively. The performance of the Andromeda (MaxQuant) search engine was intermediate with 83 identified peptides. Apart from the expected whey protein-derived sequences, several peptides were derived from co-purified caseins. The peptide lists identified with the three search engines are reported in Tables S1–S3 of the Supporting Information.

The performances of the search engines were also evaluated combining the results of both non-reduced and DTT-reduced digests analyzed with the inclusion of monocharged ions. This workflow significantly increased the list of peptides identified by the three search engines, especially for Andromeda, which was particularly effective to assign small monocharged peptide ions. However, no peptide shorter than five amino acid residues was confidently assigned by the search engines.

The Venn diagram of the peptide intersections relevant to this analysis is shown in Figure 5B, and the non-redundant peptide lists are reported in Tables S4–S6 of the Supporting

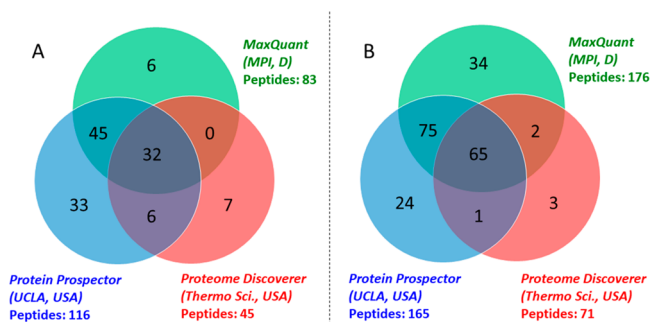


Figure 5. Outcomes of the peptide identification with MaxQuant, ProteinProspector, and Proteome Discoverer 2.1 search engines, including only multicharged ions in the m/z 1800–350 range (A) or both mono- and multicharged ions in the m/z 1600–200 range (B), merging entries obtained before and after Cys reduction. The Venn diagrams show the number of peptides identified by each of the search engines, their mutual intersection, and the intersection between the three search engines. Peptides arising from hydrolysis of the digestive enzymes were excluded from the list.

Information. Some peptides were retrieved through the MS/MS fragmentation of both mono- and double-charged ions. This information appears in the MaxQuant report (Table S5 of the Supporting Information), although it is valid also for the other search engines. Allowing the MS/MS fragmentation of monocharged ions in the absence of the exclusion list for interfering compounds had detrimental effects on the peptide assignment (not shown). As a general trend, the peptides identified by all three search engines belonged for the most part to a few protein domains intrinsically stable to digestion.¹⁹ After reduction with DTT, several peptides containing free cysteines were identified. As a result of the incomplete characterization of the digests, it is not possible to establish whether the missed identification of peptides from specific protein domains is a consequence of their complete degradation (e.g., into free amino acids) or their survival as large peptides.

The use of single search engines provokes loss of information, because each of them considers only part of the data generated in a HPLC–MS/MS analysis as a result of the intrinsic characteristics of the algorithms. This aspect contributes to limit the ordinary characterization of complex digests. A detailed discussion of the algorithm-related performance goes beyond the purpose of this paper. However, strengths and weaknesses of the most commonly used search engines in proteomics have been evaluated and compared in specialized reviews.^{6,7}

The use of combined results from multiple search engines could enable an extended identification of MS data set,²⁹ but at least for the analysis of digestomes, this approach is not routinely employed.

These findings highlight the need of standardized peptidomic protocols for the identification of functional peptides, which are currently under development.³⁰

Identification of Disulfide-Linked Cross-Linked Peptides. Major native bovine whey proteins (including α -La, β -Lg, and BSA) contain disulfide-linked cysteines, which, upon digestion, are likely to generate cross-linked peptide heterodimers or even concatenated multimers in the case of α -La. In general, the identification of disulfide-bridged heterodimers is not an easy task as a result of no ordinary fragmentation

pattern, the superimposition of fragmentation series potentially arising from multiple peptide chains, and a still ineffective scoring system. Software tools for the identification of cross-linked peptides do exist, but they often provide scarcely reliable outcomes, also because the systematic study of the fragmentation of disulfide-linked peptides has been undertaken only very recently.³¹

We attempted the identification of cross-linked peptides using the apposite tool of the Batch-Tag web search engine (<http://prospector2.ucsf.edu>). Many of the software outcomes appeared inconsistent, while a few assignments were validated by manual inspection of MS/MS and allowed for the identification of actual cross-linked heterodimeric β -Lg peptides. For example, the MS/MS spectrum of $[M + 3H]^{3+} = 831.05$, assigned by the software to the β -Lg f(119–135)–S–S– β -Lg f(105–108) Met-oxidized heterodimer ($[M + H]^+ = 2491.15$), is shown in Figure 6. Practically, all of the main

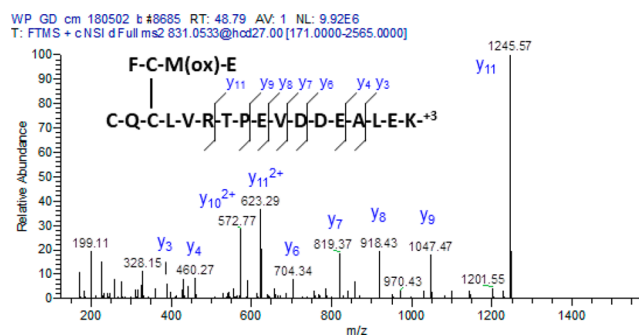


Figure 6. MS/MS spectrum of the triple-ion charge signal $[M + 3H]^{3+} = 831.05$, identified as the β -Lg f(119–135)–S–S– β -Lg f(105–108) Met-oxidized heterodimer ($[M + H]^+ = 2491.15$). The main ion fragments are assigned to the b or y series, considering β -Lg f(119–135) as the main peptide chain. From the spectrum, it cannot be established univocally if β -Lg f(105–108) Met-oxidized is linked to Cys1 or Cys2 of β -Lg f(119–135).

signals matched those expected for the fragmentation along the longest peptide chain, namely, β -Lg f(119–135). On the other hand, it was not possible to detect β -Lg f(119–135) by HPLC–MS/MS released upon reduction of the S–S bridges with DTT, even when digests were reanalyzed after alkylation (with excess of iodoacetamide) of free thiol groups.

The TIC chromatograms obtained from the GD digests before and after reduction with DTT are compared in panels A and B of Figure 3. The main chromatographic peaks appeared unaffected by reduction. On the other hand, many minor Cys-containing peptides, such as β -Lg f(155–162) and β -Lg f(61–66), with the latter deriving from both of the genetic variants β -Lg B and β -Lg A, and related shortened fragments were sequenced by MS/MS after reduction with DTT. However, it was not possible to pair them with other Cys-containing species, to match possible cross-linked peptides detected in the digests of native whey proteins. The assignment of cross-linked peptides of α -La was even more challenging to the possible occurrence of concatenated multimers, which had been previously detected in GI digests of whey proteins, by matrix-assisted laser desorption/ionization time-of-flight (MALDI–TOF) MS.¹⁵ These results highlight the complexity of the system and the difficulty to characterize the S–S-linked peptides.

Detection of Large-Sized Peptides. Large-sized peptides, also including disulfide cross-linked species, were

previously detected in an *in vitro* digest of whey proteins using MALDI–TOF MS analysis.¹⁵ We specifically searched the MW signals previously observed in the HPLC–MS/MS analysis obtained in this study, without detecting any of them. The HPLC–MS/MS analysis of the 12% TCA-insoluble fraction of the digest enhanced the detection of relatively high-molecular-weight polypeptides, also including many components with MW ranging between 8 and 13.9 kDa, which were not unequivocally assigned. These compounds also occurred in the HPLC–MS/MS analysis of the unfractionated digests, albeit at a very low relative intensity. For instance, Figure 3C shows the MS spectrum at $t_r = 98.1$ – 100.4 min of the HPLC–MS/MS run of the unfractionated sample, which contained multicharged signals of a polypeptide with MW of 13 954. This specific component was no longer detected in the HPLC–MS/MS analysis after reduction with DTT.

In agreement with the RP-HPLC analysis, the great part of these high-molecular-weight components did not occur in the HPLC–MS/MS analysis of control GI fluids, incubated in parallel during digestion of whey proteins (not shown).

Identification of Peptides Originating from Digestive Enzymes. Taking into account that the digestive enzymes used in the simulated digestion were of porcine origin, we also performed a peptide search taxonomically restricted to pig (*Sus scrofa*) proteins with the Batch-Tag web search engine. Overall, 335 peptides arising from 14 porcine proteins were identified at a high confidence degree (FDR of 1%), among which prevailed proteolytic products of pancreatic enzymes, including pancreatic α -amylase, trypsin, lipase, and chymotrypsin-like elastase isoforms. Along with a few peptides from other minor gene products, two peptides of gastric pepsin were identified as well. The presence of peptides deriving from hydrolysis of digestive enzymes increases the complexity of a digestome, introducing an additional challenge in the process of analytical characterization. The list of these peptides is reported in Table S7 of the Supporting Information and could serve as a reference for future investigations.

Challenges for the Comprehensive Identification of Digestomes. The results of this study demonstrate that several protein regions surviving the GI digestion at a low extent may remain undetected, even using sophisticated and multiplexed analytical methods.

Nutritional relevant peptides, including bioactive peptides or immunogenic/allergenic sequences, cover a broad range of MW as well as a broad range of physical–chemical properties. Within the peptidomic inventories of food digestomes, small peptides (<6–7 residues), especially the very short peptides (2–5 residues), are often underrepresented. These peptides lie at the boundary between polypeptides and small metabolites, and the analytical methods would require specific adjustments to catalogue them comprehensively. Similarly, large polypeptides (>4.5 kDa) as well as disulfide-linked dimers/multimers can be detected at relatively low sensitivity and fragmented with low efficiency by MS/MS. Even peptides with an optimal size for MS detection could escape identification as a result of the limited dynamic range of MS or an intrinsically scarce ionization efficiency.

Therefore, in general, the digestomes arising from simulated or physiological digestion of foods still appear as incomplete pictures.

As a consequence, some of the outcomes of biological assays carried out with digests of food proteins might be biased by

undisclosed species within the multitude of peptides generated by a process of (simulated) digestion.

This aspect is particularly relevant when pools of possible bioactive peptides are tested in biological assays, because the effect of bioactive peptides could be masked, attenuated, or synergized by “hidden” co-occurring components.¹³

Strikingly, the possible presence of potential immunologically active large peptides in the digests of food allergens introduces new implications in the debate about the still uncertain relationship between stability to digestion and allergenicity. Some proteins believed to be completely digested could actually produce large fragments, albeit at a very low amount, which might be, in principle, responsible for sensitization or elicitation of allergenic responses through the GI tract.

In general, the unidentified peptides of a protein digest include identifiable and unidentifiable peptides.³² The instrumental advance can contribute to minimize the number of unidentifiable peptides, whereas bioinformatics can progressively improve the assignment of identifiable peptides. Expectedly, the forthcoming research efforts will evolve in both directions in a very near future to achieve the complete characterization of complex digestomes and to bridge the gaps still existing between composition, structural properties, and biological effects of peptides originating from dietary proteins.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jafc.9b02342.

List of milk-derived peptides identified with the three different search engines, with HPLC–MS/MS analysis in the m/z 1800–350 range, allowing for fragmentation of multicharged ions (Tables S1–S3), list of milk-derived peptides identified with the three different search engines, with HPLC analysis in the m/z 1600–200 range, allowing fragmentation of mono- and multicharged ions with an exclusion list for interfering components (Tables S4–S6), and list of peptides deriving from digestive enzymes identified by HPLC–MS/MS (Table S7) (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*Telephone: +39-0825-299521. E-mail: picariello@isa.cnr.it.

ORCID

Gianfranco Mamone: 0000-0002-0394-7617

Gianluca Picariello: 0000-0003-1290-1239

Notes

The authors declare no competing financial interest.

■ REFERENCES

- (1) Daliri, E. B.-M.; Oh, D. H.; Lee, B. H. Bioactive Peptides. *Foods* **2017**, *6* (5), 32.
- (2) Agyei, D.; Tsopmo, A.; Udenigwe, C. C. Bioinformatics and peptidomics approaches to the discovery and analysis of food-derived bioactive peptides. *Anal. Bioanal. Chem.* **2018**, *410*, 3463–3472.
- (3) Lahrichi, S. L.; Affolter, M.; Zolezzi, I. S.; Panchaud, A. Food peptidomics: Large scale analysis of small bioactive peptides—a pilot study. *J. Proteomics* **2013**, *88*, 83–91.
- (4) Dallas, D. C.; Guerrero, A.; Parker, E. A.; Robinson, R. C.; Gan, J.; German, J. B.; Barile, D.; Lebrilla, C. B. Current peptidomics:

Applications, purification, identification, quantification, and functional analysis. *Proteomics* **2015**, *15* (5–6), 1026–1038.

(5) Giacometti, J.; Buretić-Tomljanović, A. Peptidomics as a tool for characterizing bioactive milk peptides. *Food Chem.* **2017**, *230*, 91–98.

(6) Verheggen, K.; Raeder, H.; Berven, F. S.; Martens, L.; Barsnes, H.; Vaudel, M. Anatomy and evolution of database search engines—A central component of mass spectrometry based proteomic workflows. *Mass Spectrom. Rev.* **2017**, DOI: [10.1002/mas.21543](https://doi.org/10.1002/mas.21543).

(7) Nesvizhskii, A. I. A survey of computational methods and error rate estimation procedures for peptide and protein identification in shotgun proteomics. *J. Proteomics* **2010**, *73*, 2092–2123.

(8) Le Maux, S.; Nongonierma, A. B.; FitzGerald, R. J. Improved short peptide identification using HILIC–MS/MS: Retention time prediction model based on the impact of amino acid position in the peptide sequence. *Food Chem.* **2015**, *173*, 847–854.

(9) Lahrichi, S. L.; Affolter, M.; Zolezzi, I. S.; Panchaud, A. Food peptidomics: Large scale analysis of small bioactive peptides—A pilot study. *J. Proteomics* **2013**, *88*, 83–91.

(10) Herregods, G.; Van Camp, J.; Morel, N.; Ghesquière, B.; Gevaert, K.; Vercruyse, L.; Dierckx, S.; Quanten, E.; Smagghe, G. Angiotensin I-converting enzyme inhibitory activity of gelatin hydrolysates and identification of bioactive peptides. *J. Agric. Food Chem.* **2011**, *59*, 552–558.

(11) O’Keeffe, M. B.; FitzGerald, R. J. Identification of short peptide sequences in complex milk protein hydrolysates. *Food Chem.* **2015**, *184*, 140–146.

(12) Wang, H.; Bennett, P. High-Resolution Accurate Mass Orbitrap and Its Application in Protein Therapeutics Bioanalysis. In *Protein Analysis using Mass Spectrometry: Accelerating Protein Biotherapeutics from Lab to Patient*; Lee, M. S., Ji, Q. C., Eds.; John Wiley & Sons, Inc.: Hoboken, NJ, 2017; Chapter 2, pp 11–20, DOI: [10.1002/9781119371779.ch2](https://doi.org/10.1002/9781119371779.ch2).

(13) Picariello, G.; Ferranti, P.; Fierro, O.; Mamone, G.; Caira, S.; Di Luccia, A.; Monica, S.; Addeo, F. Peptides surviving the simulated gastrointestinal digestion of milk proteins: Biological and toxicological implications. *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.* **2010**, *878*, 295–308.

(14) Pieri, M.; Lombardi, A.; Basilicata, P.; Mamone, G.; Picariello, G. Proteomics in Forensic Sciences: Identification of the Nature of the Last Meal at Autopsy. *J. Proteome Res.* **2018**, *17*, 2412–2420.

(15) Pekar, J.; Ret, D.; Untersmayr, E. Stability of allergens. *Mol. Immunol.* **2018**, *100*, 14–20.

(16) Shan, L.; Molberg, Ø.; Parrot, I.; Hausch, F.; Filiz, F.; Gray, G. M.; Sollid, L. M.; Khosla, C. Structural basis for gluten intolerance in celiac sprue. *Science* **2002**, *297* (5590), 2275–2279.

(17) Mamone, G.; Ferranti, P.; Rossi, M.; Roepstorff, P.; Fierro, O.; Malorni, A.; Addeo, F. Identification of a peptide from alpha-gliadin resistant to digestive enzymes: Implications for celiac disease. *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.* **2007**, *855* (2), 236–241.

(18) Bøgh, K. L.; Madsen, C. B. Food Allergens: Is There a Correlation between Stability to Digestion and Allergenicity? *Crit. Rev. Food Sci. Nutr.* **2016**, *56* (9), 1545–1567.

(19) Picariello, G.; Miralles, B.; Mamone, G.; Sánchez-Rivera, L.; Recio, I.; Addeo, F.; Ferranti, P. Role of intestinal brush border peptidases in the simulated digestion of milk proteins. *Mol. Nutr. Food Res.* **2015**, *59* (5), 948–956.

(20) Akkerdaas, J.; Totis, M.; Barnett, B.; Bell, E.; Davis, T.; Edrington, T.; Glenn, K.; Graser, G.; Herman, R.; Knulst, A.; Ladics, G.; McClain, S.; Poulsen, L. K.; Ranjan, R.; Rasle, J. B.; Serrano, H.; Speijer, D.; Wang, R.; Pereira Mouries, L.; Capt, A.; van Ree, R. Protease resistance of food proteins: A mixed picture for predicting allergenicity but a useful tool for assessing exposure. *Clin. Transl. Allergy* **2018**, *8*, 30.

(21) Picariello, G.; Mamone, G.; Nitride, C.; Addeo, F.; Ferranti, P. Protein digestomics: Integrated platforms to study food protein digestion and derived functional and active peptides. *TrAC, Trends Anal. Chem.* **2013**, *52*, 120–134.

(22) Minekus, M.; Alming, M.; Alvito, P.; Ballance, S.; Bohn, T.; Bourlieu, C.; Carrière, F.; Boutrou, R.; Corredig, M.; Dupont, D.; Dufour, C.; Egger, L.; Golding, M.; Karakaya, S.; Kirkhus, B.; Le Feunteun, S.; Lesmes, U.; Macierzanka, A.; Mackie, A.; Marze, S.; McClements, D. J.; Ménard, O.; Recio, I.; Santos, C. N.; Singh, R. P.; Vegarud, G. E.; Wickham, M. S.; Weitschies, W.; Brodtkorb, A. A standardised static in vitro digestion method suitable for food—An international consensus. *Food Funct.* **2014**, *5*, 1113–1124.

(23) Ferranti, P.; Mamone, G.; Picariello, G.; Addeo, F. The “dark side” of β -lactoglobulin: Unedited structural features suggest unexpected functions. *J. Chromatogr. A* **2011**, *1218*, 3423–3431.

(24) Picariello, G.; De Cicco, M.; Nocerino, R.; Paparo, L.; Mamone, G.; Addeo, F.; Berni Canani, R. Excretion of Dietary Cow’s Milk Derived Peptides Into Breast Milk. *Front. Nutr.* **2019**, *6*, 25.

(25) Reddy, I. M.; Kella, N. K. D.; Kinsella, J. E. Structural and conformational basis of the resistance of β -lactoglobulin to peptic and chymotryptic digestion. *J. Agric. Food Chem.* **1988**, *36*, 737–741.

(26) Tsutsumi, R.; Tsutsumi, Y. M. Peptides and Proteins in Whey and Their Benefits for Human Health. *Austin J. Nutr. Food Sci.* **2014**, *1*, 1002.

(27) Picariello, G.; Addeo, F.; Ferranti, P.; Nocerino, R.; Paparo, L.; Passariello, A.; Dallas, D. C.; Robinson, R. C.; Barile, D.; Canani, R. B. Antibody-independent identification of bovine milk-derived peptides in breast-milk. *Food Funct.* **2016**, *7*, 3402–3409.

(28) Nongonierma, A. B.; FitzGerald, R. J. Strategies for the discovery and identification of food protein-derived biologically active peptides. *Trends Food Sci. Technol.* **2017**, *69*, 289–305.

(29) Shteynberg, D.; Nesvizhskii, A. I.; Moritz, R. L.; Deutsch, E. W. Combining results of multiple search engines in proteomics. *Mol. Cell. Proteomics* **2013**, *12*, 2383–2393.

(30) Dallas, D.; Nielsen, S. D. Milk peptidomics to identify functional peptides and for quality control of dairy products. In *Peptidomics: Methods and Strategies*; Schrader, M.; Fricker, L., Eds.; Springer: New York, 2018; Vol. 1719, pp 223–240, DOI: [10.1007/978-1-4939-7537-2_15](https://doi.org/10.1007/978-1-4939-7537-2_15).

(31) Liang, Z.; McGuinness, K. N.; Crespo, A.; Zhong, W. Characterization of Disulfide-Linked Peptides Using Tandem Mass Spectrometry Coupled with Automated Data Analysis Software. *J. Am. Soc. Mass Spectrom.* **2018**, *29*, 903–912.

(32) Mancuso, F. M.; Cappadona, S.; Sabidó, E. Bioinformatic Approaches to Increase Proteome Coverage. In *Comprehensive Analytical Chemistry*; Simó, C., Cifuentes, A., García-Cañas, V., Eds.; Elsevier: Netherlands, 2014; Comprehensive Analytical Chemistry, Fundamentals of Advanced Omics Technologies: From Genes to Metabolites, Vol. 63, Chapter 16, pp 385–419, DOI: [10.1016/B978-0-444-62651-6.00017-9](https://doi.org/10.1016/B978-0-444-62651-6.00017-9).